Appendix D: Statistics Used on Item Labels

Appendix C: Statistics Used on Item Labels

CID Company identification number for the item.

Maturity Function of the reviewed item.

Form Form numbers that contain the reviewed item.

Position Position numbers in the test for the reviewed item (given for each form that the item appears on).

Type Item type: MC – multiple-choice item, CR – constructed-response item, WR – writing.

Key The correct answer for an MC item.

Max The maximum score point for a CR or a writing item.

P-value The percent of students who answered the item correctly. Its theoretical range is 0-1. It indicates item difficulty. Items with high p-values, such as .90, are relatively easy items. Those with p-values below .50 are relatively difficult items. P-values depend on the group of examinees who take the test.

Adj. P value Computed by dividing the item mean by the score range. It is equivalent to the p-value for the MC items when the score point is awarded either 1 or 0.

N-count The number of tested students who were administered the item.

Rasch Difficulty The usual range of Rasch difficulties is from -3 to +3 with mean of 0 and standard deviation of 1. 0 means medium difficulty. Positive values mean difficult items. Negative values mean easy items.

PB Point-biserial correlation shows the relationship between a student's performance

Correlation on the item and performance on the test as a whole. A high point-biserial correlation (e.g., above .50) indicates that students who answered the item correctly on the item achieved higher total scores on the test than those who answered the item incorrectly on the item. Values less than .25 may indicate a weaker than desired relationship. Note that extremely difficult or extremely easy items may have point-biserial correlation artificially reduced.

Item-Total Item-total correlation shows the relationship between a student's performance on

Corr.	the item and performance on the test as a whole. A high item-total correlation (e.g., above .50) indicates that students who earned more points on the item achieved higher total scores on the test than those who earned fewer points on the item. Values less than .25 may indicate a weaker than desired relationship. Note that extremely difficult or extremely easy items may have item-total correlation artificially reduced.
FIT Flag	This flag indicates that two fit indices are out of the desired range. It means the Item may have not misfit or overfit the measurement model specified for the test analysis.
Difficulty Flag	This flag indicates that P-value, or adjusted p-value, or Rasch difficulty is out of the desired range.
PB Correlation Flag	This flag indicates that a MC item point-biserial correlation is smaller than the desired range of larger than 0.25.
Item-Total Corr. Flag	This flag indicates that a CR or a Writing item point-biserial correlation is smaller than the desired range of larger than 0.25.
Option Quality Flag	This flag indicates that a MC item may have a key problem. It could be that the key is not correct or it was miskeyed in scoring.
Score Point Dist. Flag	This flag indicates that a CR or a Writing item may have a scoring rubric problem. It could be the sample answer for each score point was not correctly identified.
Option Analysis	Percent of students who selected options A, B, C, and D, or did not choose any option (Omit) for all students and for subgroups by gender and ethnicity.
Score Point Distribution	Percent of students who earned each valid score point and who did not answer the CR or writing item for all students and for subgroups by gender and ethnicity.
Option PB Correlation	Point-biserial correlation for each of a MC item options. The key option point-biserial correlation should be positive and high. The non-keyed option point-biserial should be negative and low.

Omit PB Correlation	Point-biserial correlation for omit of a CR or Writing item. The omit point-biserial correlation should be negative.
Invalid Codes	The codes for invalid responses for a CR or a writing item.
DIF	Differential Item Functioning index. It indicates whether the reviewed item favors a particular subgroup of the student population; thus that group of students may have a higher chance of answering the item correctly or earn higher score point than the contrasted group. The focused group is often the minority group such as female in the gender group comparison, and black in the ethnic group comparison. The reference group is often the majority group which is male in the gender group comparison, and white in the ethnic group comparison.